(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété Intellectuelle
Bureau international

(43) Date de la publication internationale
9 octobre 2003 (09.10.2003)

PCT

(10) Numéro de publication internationale
WO 03/083830 A1

(51) Classification internationale des brevets[7] :
G10L 15/14, 15/18

(21) Numéro de la demande internationale :
PCT/FR03/00653

(22) Date de dépôt international : 19 mars 2003 (19.03.2003)

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :
02/04285 29 mars 2002 (29.03.2002) FR

(71) Déposant *(pour tous les États désignés sauf US)* : **FRANCE TELECOM** [FR/FR]; 6, place d'Alleray, F-75015 Paris (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants *(pour US seulement)* : **FERRIEUX, Alexandre** [FR/FR]; 4, Hent Al Lann, F-22560 Pleumeur Bodou (FR). **DELPHIN-POULAT, Lionel** [FR/FR]; résidence Kergemar A2, F-22300 Lannion (FR).

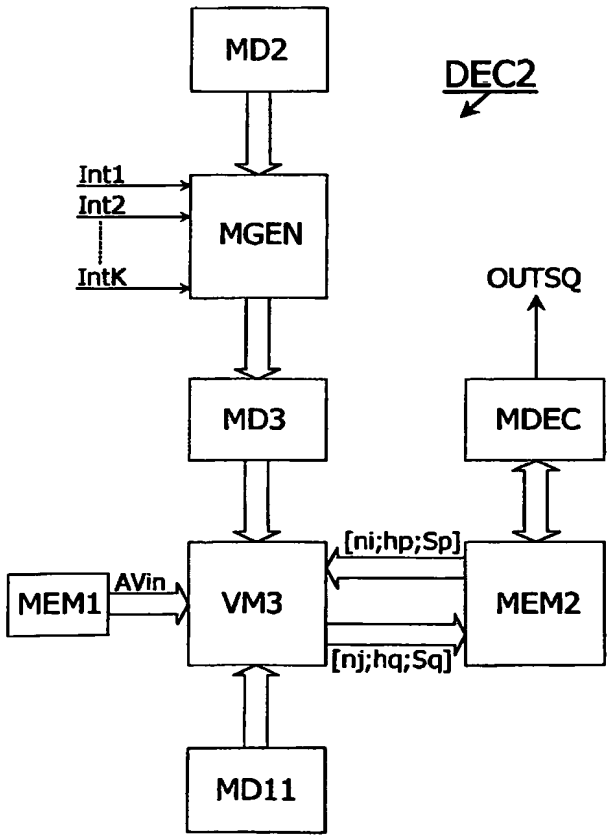
(74) Mandataire : **MAILLET, Alain**; Cabinet Le Guen Maillet, 5, place Newquay, B.P. 70250, F-35802 Dinard Cedex (FR).

(81) États désignés *(national)* : AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, RO, RU, SC, SD, SE, SG, SK,

[Suite sur la page suivante]

(54) Title: SPEECH RECOGNITION METHOD

(54) Titre : PROCEDE DE RECONNAISSANCE DE LA PAROLE



(57) **Abstract:** The invention relates to a method of translating input data AVin into at least one output sequence (OUTSQ). The inventive method comprises a decoding step during which sub-lexical entities having representative input data (AVin) are identified using a first model (MD 11) and during which different possible combinations of the aforementioned sub-lexical entities are generated as said sub-lexical entities are identified and with reference to a second model (MD3). The invention also involves the storing of several possible combinations [nj;hq;Sq] of the above-mentioned sub-lexical entities, the most likely combination being intended to form the output lexical sequence (OUTSQ) and one such storage operation enabling the structure of the second model (MD3) to be simplified.

(57) **Abrégé :** La présente invention concerne un procédé de traduction de données d'entrée AVin en au moins une séquence de sortie (OUTSQ), incluant une étape de décodage au cours de laquelle des entités sous-lexicales dont les données d'entrée (Avin) sont représentatives sont identifiées au moyen d'un premier modèle (MD 11), et au cours de laquelle sont générées, au fur et à mesure que les entités sous-lexicales sont identifiées et en référence à au moins un deuxième modèle (MD3), diverses combinaisons possibles desdites entités sous-lexicales. L'invention prévoit de mémoriser une pluralité de combinaisons possibles [nj;hq;Sq] desdites entités sous-lexicales, la combinaison la plus vraisemblable étant destinée à former la séquence lexicale de sortie (OUTSQ), une telle mémorisation permettant de simplifier la structure du deuxième modèle (MD3).

WO 03/083830 A1

SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) **États désignés** *(régional)* : brevet ARIPO (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), brevet eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Publiée :

— *avec rapport de recherche internationale*

En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.

PROCEDE DE RECONNAISSANCE DE LA PAROLE

Procédé de traduction de données autorisant une gestion de mémoire simplifiée

La présente invention concerne un procédé de traduction de données d'entrée en au moins une séquence lexicale de sortie, incluant une étape de décodage des données d'entrée au cours de laquelle des entités lexicales dont lesdites données sont représentatives sont identifiées au moyen d'au moins un modèle.

De tels procédés sont communément utilisés dans des applications de reconnaissance de parole, où au moins un modèle est mis en œuvre pour reconnaître des symboles acoustiques présents dans les données d'entrée, un symbole pouvant être constitué par exemple par un ensemble de vecteurs de paramètres d'un espace acoustique continu, ou encore par un label attribué à une entité sous-lexicale.

Dans certaines applications, le qualificatif "lexical" s'appliquera à une phrase considérée dans son ensemble, en tant que suite de mots, et les entités sous-lexicales seront alors des mots, alors que dans d'autres applications, le qualificatif "lexical" s'appliquera à un mot, et les entités sous-lexicales seront alors des phonèmes ou encore des syllabes aptes à former de tels mots, si ceux-ci sont de nature littérale, ou des chiffres, si les mots sont de nature numérique, c'est-à-dire des nombres.

Une première approche pour opérer une reconnaissance de parole consiste à utiliser un type particulier de modèle qui présente une topologie régulière et est destiné à apprendre toutes les variantes de prononciation de chaque entité lexicale, c'est-à-dire par exemple un mot, inclus dans le modèle. Selon cette première approche, les paramètres d'un ensemble de vecteurs acoustiques propre à chaque symbole d'entrée correspondant à un mot inconnu doivent être comparés à des ensembles de paramètres acoustiques correspondant chacun à l'un des très nombreux symboles contenus dans le modèle, afin d'identifier un symbole modélisé auquel correspond le plus vraisemblablement le symbole d'entrée. Une telle approche garantit en théorie un fort taux de reconnaissance si le modèle utilisé est bien conçu, c'est-à-dire quasi-exhaustif, mais une telle quasi-exhaustivité ne peut être obtenue qu'au prix d'un long processus d'apprentissage du modèle, qui doit assimiler une énorme quantité de données représentatives de toutes les variantes de prononciation de chacun des mots inclus dans ce modèle. Cet apprentissage est en principe réalisé en faisant prononcer par un grand nombre de personnes tous les mots d'un vocabulaire donné, et à enregistrer toutes les variantes de prononciation de ces mots. Il apparaît clairement que la construction d'un modèle lexical quasi-exhaustif n'est pas envisageable en pratique pour des vocabulaires présentant une taille supérieure à quelques centaines de mots.

Une deuxième approche a été conçue dans le but de réduire le temps d'apprentissage nécessaire aux applications de reconnaissance de parole, réduction qui est essentielle à des applications de traduction sur de très grands vocabulaires pouvant contenir plusieurs centaines de milliers de mots, laquelle deuxième approche consiste à opérer une factorisation des entités lexicales en les considérant comme des assemblages d'entités sous-lexicales, à générer un modèle sous-lexical modélisant lesdites entités sous-lexicales en vue de permettre leur identification dans les données d'entrée, et un modèle d'articulation modélisant différentes combinaisons possibles de ces entités sous-lexicales. Selon cette deuxième approche, un nouveau modèle dynamique formant le modèle d'articulation est constitué à partir de chaque entité sous-lexicale nouvellement identifiée dans les données d'entrée, lequel modèle

dynamique rend compte de tous les assemblages rendus possibles en partant de l'entité sous-lexicale considérée, et détermine une valeur de vraisemblance pour chaque assemblage possible.

Une telle approche, décrite par exemple au chapitre 16 du manuel "Automatic Speech and Speaker Recognition" édité par Kluwer Academic Publishers, permet de réduire considérablement, par rapport au modèle utilisé dans le cadre de la première approche décrite plus haut, les durées individuelles des processus d'apprentissage du modèle sous-lexical et du modèle d'articulation, car chacun de ces modèles présente une structure simple par rapport au modèle lexical utilisé dans la première approche.

Cependant, dans la plupart des implémentations connues de la deuxième approche décrite ci-dessus, le modèle sous-lexical est dupliqué à de multiples reprises dans le modèle d'articulation. Ceci peut être aisément compris en considérant un exemple où l'unité lexicale est une phrase et les unités sous-lexicales sont des mots. Si le modèle d'articulation est d'un type bi-gramme, c'est-à-dire qu'il rend compte de possibilités d'assemblage de deux mots successifs et de probabilités d'existence de tels assemblages, chaque mot retenu à l'issue de la sous-étape d'identification devra être étudié, en référence au modèle d'articulation, avec tous les autres mots retenus ayant pu précéder le mot considéré. Si P mots ont été retenus à l'issue de la sous-étape d'identification, P couples de mots devront être construits pour chaque mot à identifier, avec P valeurs de probabilité d'existence, chacune associée à un couple possible. Dans le cas d'un modèle d'articulation plus réaliste de type tri-gramme, qui rend compte de possibilités d'assemblage de trois mots successifs et de probabilités d'existence de tels assemblages, le modèle d'articulation devra comporter, pour chaque mot à identifier, P fois P triplets de mots avec autant de valeurs de probabilité d'existence. Les modèles d'articulation mis en œuvre dans la deuxième approche ont donc une structure simple, mais représentent un volume considérable de données à mémoriser, à mettre à jour et à consulter. On conçoit aisément que la création et l'exploitation de tels modèles donne lieu à des accès mémoire dont la gestion est rendue complexe par le volume de données à traiter, et par la répartition desdites données. Dans des applications de type langage naturel, pour lesquelles des modèles

plus réalistes de type N-gramme, où N est le plus souvent supérieur à deux, sont mis en œuvre, les accès mémoire évoqués précédemment présentent des temps d'exécution incompatibles avec des contraintes de type "temps réel" nécessitant des accès mémoire très rapides.

Par ailleurs, chaque mot peut lui-même être considéré vis-à-vis de syllabes ou de phonèmes qui le composent comme une entité lexicale d'un niveau inférieur à celui d'une phrase, entité lexicale pour la modélisation de laquelle il faut également recourir à un modèle d'articulation de type N-gramme avec plusieurs dizaines d'entités sous-lexicales possibles dans le cas des phonèmes.

Il apparaît clairement que les multiples duplications des modèles sous-lexicaux auxquelles font appel les modèles d'articulation dans les implémentations connues de la deuxième approche prohibent l'utilisation de celle-ci dans des applications de reconnaissance de parole dans le cadre d'applications de type très grands vocabulaires, qui comportent plusieurs centaines de milliers de mots.

L'invention a pour but de remédier dans une large mesure à cet inconvénient, en proposant un procédé de traduction qui ne nécessite pas de multiples duplications de modèles sous-lexicaux pour valider des assemblages d'entités sous-lexicales, et simplifie ainsi l'implémentation dudit procédé de traduction, et en particulier la gestion d'accès mémoire utiles à ce procédé.

En effet, un procédé de traduction conforme au paragraphe introductif, incluant une étape de décodage au cours de laquelle des entités sous-lexicales dont les données d'entrée sont représentatives sont identifiées au moyen d'un premier modèle construit sur la base d'entités sous-lexicales prédéterminées, et au cours de laquelle sont générées, au fur et à mesure que les entités sous-lexicales sont identifiées et en référence à au moins un deuxième modèle construit sur la base d'entités lexicales, diverses combinaisons possibles desdites entités sous-lexicales, est caractérisé selon l'invention en ce que l'étape de décodage inclut une sous-étape de mémorisation d'une pluralité de combinaisons possibles desdites entités sous-lexicales, la combinaison la plus vraisemblable étant destinée à former la séquence lexicale de sortie.

Du fait que divers assemblages d'entités sous-lexicales sont mémorisés au fur et à mesure que ces entités sont produites, il n'est plus nécessaire de construire après identification de chacune desdites entités sous-lexicales un modèle dynamique reprenant toutes les entités sous-lexicales possibles, ce qui permet d'éviter les duplications évoquées plus haut et les problèmes de gestion mémoire y afférant.

La possibilité de mémoriser plusieurs combinaisons différentes permet de garder une trace de plusieurs assemblages possibles d'entités sous-lexicales, chacun présentant une vraisemblance propre à l'instant où cet assemblage est généré, laquelle vraisemblance pouvant être affectée favorablement ou défavorablement après analyse de sous-entités lexicales ultérieurement produites. Ainsi, une sélection d'un assemblage présentant la plus forte vraisemblance à un instant donné, mais qui sera finalement jugé peu vraisemblable à la lumière d'entités sous-lexicales ultérieures ne provoquera pas une élimination systématique d'autres assemblages, qui pourront finalement s'avérer plus pertinents. Cette variante de l'invention permet donc de conserver des données représentant, sous forme de différents historiques, différentes interprétations des données d'entrée, interprétations dont la plus vraisemblable pourra être identifiée et retenue pour former la séquence lexicale de sortie lorsque toutes les entités sous-lexicales auront elles-même été identifiées.

Dans un mode de réalisation particulier de cette variante de l'invention, la mémorisation d'une combinaison est assujettie à une validation opérée en référence au moins au deuxième modèle.

Ce mode de réalisation permet de réaliser de manière simple un filtrage des assemblages qui paraissent peu vraisemblables à la lumière du deuxième modèle. Seuls seront retenus et mémorisés les assemblages les plus plausibles, les autres assemblages n'étant pas mémorisés et donc pas ultérieurement pris en considération.

Dans une variante de ce mode de réalisation, la validation de mémorisation pourra être effectuée en référence à plusieurs modèles de niveaux équivalents et/ou différentes, un niveau rendant compte de la nature sous-lexicale, lexicale ou encore grammaticale d'un modèle.

Dans un mode de réalisation particulièrement avantageux de cette variante de l'invention, une validation de mémorisation d'une combinaison est accompagnée d'une attribution à la combinaison à mémoriser d'une valeur de probabilité représentative de la vraisemblance de ladite combinaison.

Ce mode de réalisation permet de moduler la nature binaire du filtrage opérée par la validation ou l'absence de validation de la mémorisation d'une combinaison, en affectant une appréciation quantitative à chaque combinaison mémorisée. Ceci permettra une meilleure appréciation de la vraisemblance des diverses combinaisons qui auront été mémorisées, et donc une traduction de meilleure qualité des données d'entrée.

On pourra en outre prévoir que différentes opérations de validation portant sur différentes combinaisons relatives à un même état du premier modèle sont exécutées de façon contiguë dans le temps.

Ceci permettra de réduire encore le volume des accès mémoire et des duplications de calcul, en traitant en une seule fois toute une famille d'informations qu'il faudra sinon mémoriser et lire à de multiples reprises.

Dans un mode de réalisation particulier de l'invention, l'étape de décodage met en œuvre un algorithme de Viterbi appliqué à un premier modèle de Markov constitué d'entités sous-lexicales, sous contrôle dynamique d'un deuxième modèle de Markov représentatif de combinaisons possibles d'entités sous-lexicales.

Ce mode de réalisation est avantageux en ce qu'il utilise des moyens éprouvés et individuellement connus de l'homme du métier, le contrôle dynamique obtenu grâce au deuxième modèle de Markov permettant de valider les assemblages d'entités sous-lexicales au fur et à mesure que lesdites entités sont identifiées au moyen de l'algorithme de Viterbi, ce qui évite d'avoir à construire après identification de chaque entité sous-lexicale un nouveau modèle dynamique reprenant toutes les entités sous-lexicales possibles semblable à ceux utilisés dans les implémentations connues de la deuxième approche évoquée plus haut.

L'invention concerne également un système de reconnaissance de signaux acoustiques mettant en œuvre un procédé tel que décrit ci-dessus.

Les caractéristiques de l'invention mentionnées ci-dessus, ainsi que d'autres, apparaîtront plus clairement à la lecture de la description suivante d'un exemple de réalisation, ladite description étant faite en relation avec les dessins joints, parmi lesquels :

La Fig.1 est un schéma fonctionnel décrivant un système de reconnaissance acoustique dans lequel un procédé conforme à l'invention est mis en œuvre,

La Fig.2 est un schéma fonctionnel décrivant un décodeur destiné à exécuter une première étape de décodage dans ce mode de mise en œuvre particulier de l'invention, et

La Fig.3 est un schéma fonctionnel décrivant un décodeur destiné à exécuter une deuxième étape de décodage conforme au procédé selon l'invention.

La Fig.1 représente schématiquement un système SYST de reconnaissance acoustique selon un mode de mise en œuvre particulier de l'invention, destiné à traduire un signal acoustique d'entrée ASin en une séquence lexicale de sortie OUTSQ. Le signal d'entrée ASin est constitué par un signal électronique analogique, qui pourra provenir par exemple d'un microphone non représenté sur la figure. Dans le mode de réalisation décrit ici, le système SYST inclut un étage d'entrée FE, contenant un dispositif de conversion analogique/numérique ADC, destiné à fournir un signal numérique ASin(1:n), formé d'échantillons ASin(1), ASin(2)...ASin(n) codés chacun sur b bits, et représentatif du signal acoustique d'entrée ASin, et un module d'échantillonnage SA, destiné à convertir le signal acoustique numérisé ASin(1:n) en une séquence de vecteurs acoustiques AVin, chaque vecteur étant muni de composantes AV1, AV2...AVr où r est la dimension d'un espace acoustique défini pour une application donnée à laquelle le système de traduction SYST est destiné, chacune des composantes AVi (pour i=1 à r) étant évaluée en fonction de caractéristiques propres à cet espace acoustique.

Le système SYST inclut en outre un premier décodeur DEC1, destiné à fournir une sélection Int1, Int2...IntK d'interprétations possibles de la séquence de vecteurs acoustiques AVin en référence à un modèle MD1 construit sur la base d'entités sous-lexicales prédéterminées.

Le système SYST inclut de plus un deuxième décodeur DEC2 dans lequel un procédé de traduction conforme à l'invention est mis en œuvre en vue d'analyser des données d'entrée constituées par les vecteurs acoustiques AVin en référence à un premier modèle construit sur la base d'entités sous-lexicales prédéterminées, par exemple le modèle MD1, et en référence à au moins un deuxième modèle MD2 construit sur la base d'entités lexicales représentatives des interprétations Int1, Int2...IntK sélectionnées par le premier décodeur DEC1, en vue d'identifier celle desdites interprétations qui devra constituer la séquence lexicale de sortie OUTSQ.

La fig.2 représente plus en détail le premier décodeur DEC1, qui inclut une première machine de Viterbi VM1, destinée à exécuter une première sous-étape de décodage de la séquence de vecteurs acoustiques AVin représentative du signal acoustique d'entrée et préalablement générée par l'étage d'entrée FE, laquelle séquence sera en outre avantageusement mémorisée dans une unité de stockage MEM1 pour des raisons qui apparaîtront dans la suite de l'exposé. La première sous-étape de décodage est opérée en référence à un modèle de Markov MD11 autorisant en boucle toutes les entités sous-lexicales, de préférence tous les phonèmes de la langue dans laquelle le signal acoustique d'entée doit être traduit si l'on considère que les entités lexicales sont des mots, les entités sous-lexicales étant représentées sous forme de vecteurs acoustiques prédéterminés.

La première machine de Viterbi VM1 est apte à restituer une séquence de phonèmes Phsq qui constitue la plus proche traduction phonétique de la séquence de vecteurs acoustiques AVin. Les traitements ultérieurs réalisés par le premier décodeur DEC1 se feront ainsi au niveau phonétique, et non plus au niveau vectoriel, ce qui réduit considérablement la complexité desdits traitements, chaque vecteur étant une entité multidimensionnelle présentant r composantes, tandis qu'un phonème peut en principe être identifié par un label unidimensionnel qui lui est propre, comme par exemple un label "OU" attribué à une voyelle orale "u", ou un label "CH" attribué à une consonne frictive non-voisée "ʃ". La séquence de phonèmes Phsq générée par la première machine de Viterbi VM1 est ainsi constituée d'une succession de labels plus aisément manipulables que ne le seraient des vecteurs acoustiques.

Le premier décodeur DEC1 inclut une deuxième machine de Viterbi VM2 destinée à exécuter une deuxième sous-étape de décodage de la séquence de phonèmes Phsq générée par la première machine de Viterbi VM1. Cette deuxième étape de décodage est opérée en référence à un modèle de Markov MD12 constitué de transcriptions sous-lexicales d'entités lexicales, c'est-à-dire dans cet exemple de transcriptions phonétiques de mots présents dans le vocabulaire de la langue dans laquelle le signal acoustique d'entrée doit être traduit. La deuxième machine de Viterbi est destinée à interpréter la séquence de phonèmes Phsq, qui est fortement bruitée du fait que le modèle MD11 utilisé par la première machine de Viterbi VM1 est d'une grande simplicité, et met en œuvre des prédictions et des comparaisons entre des suites de labels de phonèmes contenus dans la séquence de phonèmes Phsq et diverses combinaisons possibles de labels de phonèmes prévues dans le modèle de Markov MD12. Bien qu'une machine de Viterbi ne restitue usuellement que celle des séquences qui présente la plus grande probabilité, la deuxième machine de Viterbi VM2 mise en œuvre ici restituera avantageusement toutes les séquences de phonèmes lsq1, lsq2...lsqN que ladite deuxième machine VM2 aura pu réconstituer, avec des valeurs de probabilité associées p1, p2...pN qui auront été calculées pour lesdites séquences et seront représentatives de la fiabilité des interprétations du signal acoustique que ces séquences représentent.

Toutes les interprétations possibles lsq1, lsq2...lsqN étant rendues automatiquement disponibles à l'issue de la deuxième sous-étape de décodage, une sélection de K interprétations Int1, Int2...IntK qui présentent les plus fortes valeurs de probabilité est aisée quelle que soit la valeur de K qui aura été choisie.

Les première et deuxième machines de Viterbi VM1 et VM2 peuvent fonctionner en parallèle, la première machine de Viterbi VM1 générant alors au fur et à mesure des labels de phonèmes qui seront immédiatement pris en compte par la deuxième machine de Viterbi VM2, ce qui permet de réduire le délai total perçu par un utilisateur du système nécessaire à la combinaison des première et deuxième sous-étapes de décodage en autorisant la mise en œuvre de l'ensemble des ressources de calcul nécessaires au fonctionnement du premier décodeur DEC1 dès que les vecteurs

acoustiques AVin représentatifs du signal acoustique d'entrée apparaissent, et non pas après qu'ils aient été entièrement traduits en une séquence complète de phonèmes Phsq par la première machine de Viterbi VM1.

La Fig.3 représente plus en détail un deuxième décodeur DEC2 conforme à un mode de réalisation particulier de l'invention. Ce deuxième décodeur DEC2 inclut une troisième machine de Viterbi VM3 destinée à analyser la séquence de vecteurs acoustiques AVin représentative du signal acoustique d'entrée préalablement mémorisée dans l'unité de stockage MEM1.

A cet effet, la troisième machine de Viterbi VM3 est destinée à exécuter une sous-étape d'identification au cours de laquelle les entités sous-lexicales dont les vecteurs acoustiques AVin sont représentatifs sont identifiées au moyen d'un premier modèle construit sur la base d'entités sous-lexicales prédéterminées, dans cet exemple le modèle de Markov MD11 mis en œuvre dans le premier décodeur et déjà décrit plus haut.

La troisième machine de Viterbi VM3 génère en outre, au fur et à mesure que ces entités sont identifiées et en référence à au moins un modèle de Markov spécifique MD3 construit sur la base d'entités lexicales, diverses combinaisons possibles des entités sous-lexicales, la combinaison la plus vraisemblable étant destinée à former la séquence lexicale de sortie OUTSQ. Le modèle de Markov spécifique MD3 est ici spécialement généré à cet effet par un module de création de modèle MGEN, et est uniquement représentatif d'assemblages possibles de phonèmes au sein des séquences de mots formées par les diverses interprétations phonétiques Int1, Int2,...IntK du signal acoustique d'entrée délivrées par le premier décodeur, lesquels assemblages sont représentés par des sous-modèles extraits du modèle lexical MD2 par le module de création de modèle MGEN. Le modèle de Markov spécifique MD3 présente donc une taille restreinte du fait de sa spécificité.

Lorsque la troisième machine de Viterbi VM3 se trouve dans un état ni donné, auquel sont associés un historique hp et une valeur de probabilité Sp, s'il existe dans le modèle de Markov MD11 une transition dudit état ni vers un état nj munie d'un marqueur M, lequel marqueur pouvant par exemple être constitué par le label d'un

phonème dont le dernier état est ni ou d'un phonème dont le premier état est nj, la troisième machine de Viterbi VM3 associera à l'état nj un nouvel historique hq et une nouvelle valeur de probabilité Sq qui seront générés en référence au modèle spécifique MD3, sur la base de l'historique hp, de sa valeur de probabilité associée Sp et du marqueur M, la valeur de probabilité Sp pouvant en outre être également modifiée en référence au modèle de Markov MD11. Cette opération sera répétée pour tous les historiques associés à l'état ni. Si un même historique hk est associé à plusieurs reprises à un même état du modèle de Markov MD11 avec différentes valeurs de probabilité Sp1,...Spq, conformément à l'algorithme de Viterbi, seule la valeur de probabilité la plus élevée sera conservée et attribuée en tant que valeur de probabilité Sp à l'historique hk.

Chaque état nj est mémorisé dans une unité de stockage MEM2 avec ses différents historiques hq et une valeur de probabilité Sq propre à chaque historique, et ce jusqu'à ce que la troisième machine de Viterbi VM3 ait identifié tous les phonèmes contenus dans la séquence de vecteurs acoustiques d'entrée AVin et ait atteint un dernier état nf au fil d'une pluralité d'historiques hf représentant les diverses combinaisons possibles des phonèmes identifiés. Celui de ces historiques auquel aura été attribuée la plus forte valeur de probabilité Sf_{max} sera retenu par un décodeur de mémoire MDEC pour former la séquence lexicale de sortie OUTSQ.

Le modèle de Markov MD3 opère donc un contrôle dynamique permettant de valider les assemblages de phonèmes au fur et à mesure que lesdits phonèmes sont identifiés par la troisième machine de Viterbi VM3, ce qui évite d'avoir à dupliquer ces phonèmes pour former des modèles tels ceux utilisés dans les implémentations connues de la deuxième approche évoquée plus haut. De la sorte, les accès aux unités de stockage MEM1 et MEM2, ainsi qu'au différents modèles de Markov MD11, MD12, MD2 et MD3 mis en œuvre dans l'exemple décrit ci-dessus nécessitent une gestion peu complexe, du fait de la simplicité de structure desdits modèles et des informations destinées à être mémorisées et lues dans lesdites unités de stockage. Ces accès mémoire peuvent donc être exécutés suffisamment rapidement pour rendre le

système décrit dans cet exemple apte à accomplir des traductions en temps réel de données acoustiques d'entrée en séquences lexicales de sortie.

Bien que l'invention ait été décrite ici dans le cadre d'une application au sein d'un système incluant deux décodeurs disposés en cascade, il est tout-à-fait envisageable, dans d'autres modes de mise en œuvre de l'invention, de n'utiliser qu'un unique décodeur semblable au deuxième décodeur décrit plus haut, qui pourra par exemple opérer une analyse acoustico-phonétique et mémoriser, au fur et à mesure que des phonèmes seront identifiés, diverses combinaisons possibles desdits phonèmes, la combinaison de phonèmes la plus vraisemblable étant destinée à former la séquence lexicale de sortie.

REVENDICATIONS

1) Procédé de traduction de données d'entrée en au moins une séquence lexicale de sortie, incluant une étape de décodage des données d'entrée au cours de laquelle des entités sous-lexicales dont lesdites données sont représentatives sont identifiées au moyen d'un premier modèle construit sur la base d'entités sous-lexicales prédéterminées, et au cours de laquelle sont générées, au fur et à mesure que les entités sous-lexicales sont identifiées et en référence à au moins un deuxième modèle construit sur la base d'entités lexicales, diverses combinaisons possibles desdites entités sous-lexicales,

procédé caractérisé en ce que l'étape de décodage inclut une sous-étape de mémorisation d'une pluralité de combinaisons possibles desdites entités sous-lexicales, la combinaison la plus vraisemblable étant destinée à former la séquence lexicale de sortie.

2) Procédé de traduction selon la revendication 1, caractérisé en ce que la mémorisation d'une combinaison est assujettie à une validation opérée en référence au moins au deuxième modèle.

3) Procédé de traduction selon la revendication 2, caractérisé en ce qu'une validation de mémorisation d'une combinaison est accompagnée d'une attribution à la combinaison à mémoriser d'une valeur de probabilité représentative de la vraisemblance de ladite combinaison.

4) Procédé de traduction selon l'une des revendications 2 ou 3, caractérisé en ce que différentes opérations de validation portant sur différentes combinaisons relatives à un même état du premier modèle sont exécutées de façon contiguë dans le temps.

5) Procédé de traduction selon la revendication 1, caractérisé en ce que l'étape de décodage met en œuvre un algorithme de Viterbi appliqué à un premier modèle de Markov constitué d'entités sous-lexicales, sous contrôle dynamique d'un deuxième modèle de Markov représentatif de combinaisons possibles d'entités sous-lexicales.

6) Système de reconnaissance vocale mettant en œuvre un procédé de traduction conforme à l'une des revendications 1 à 5.

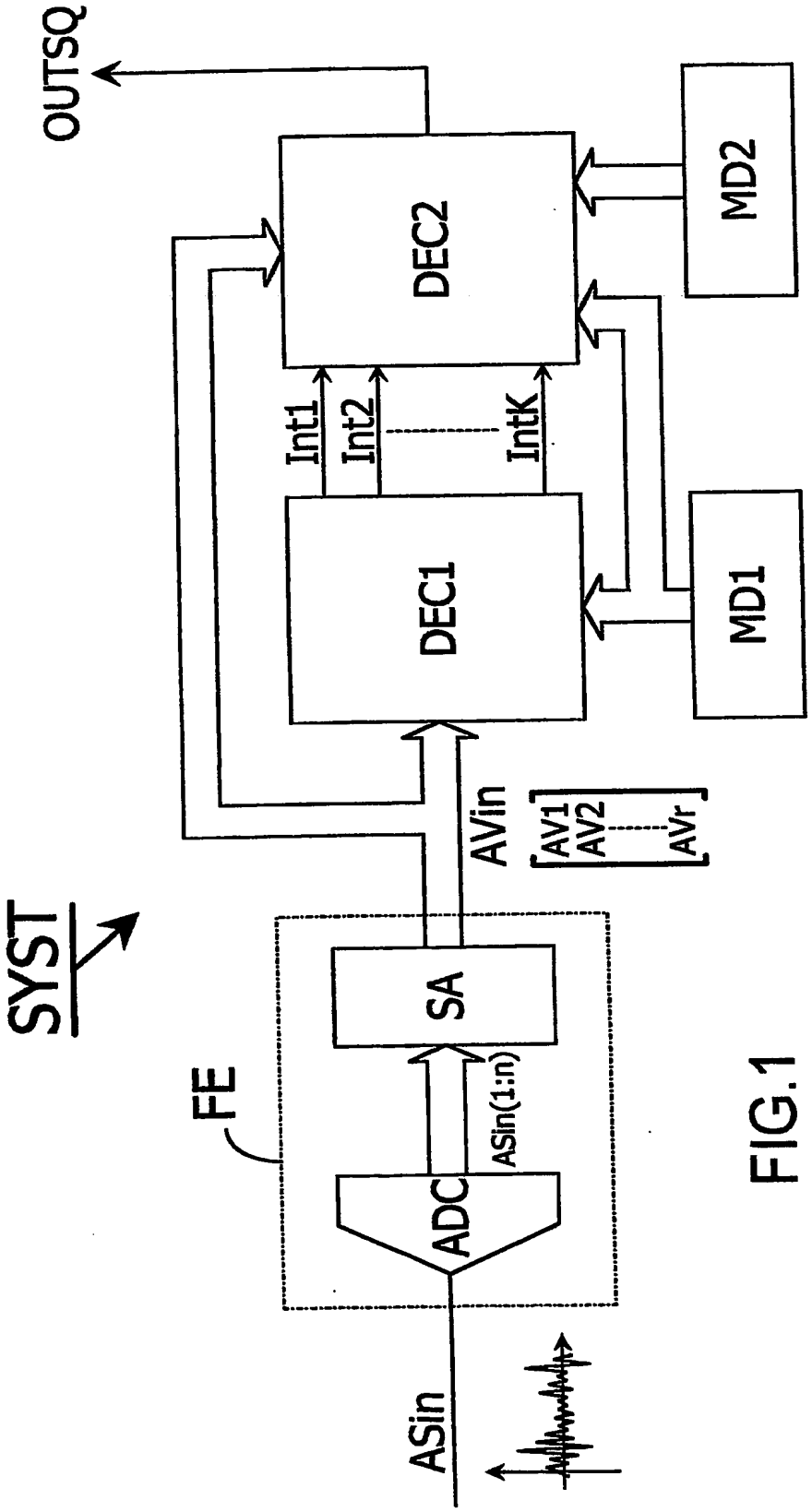

OUTSQ
DEC2
MD2
Int1
Int2
IntK
DEC1
MD1
AVin
AV1
AV2
AVr
SYST
SA
FE
ASin(1:n)
ADC
ASin

FIG.1

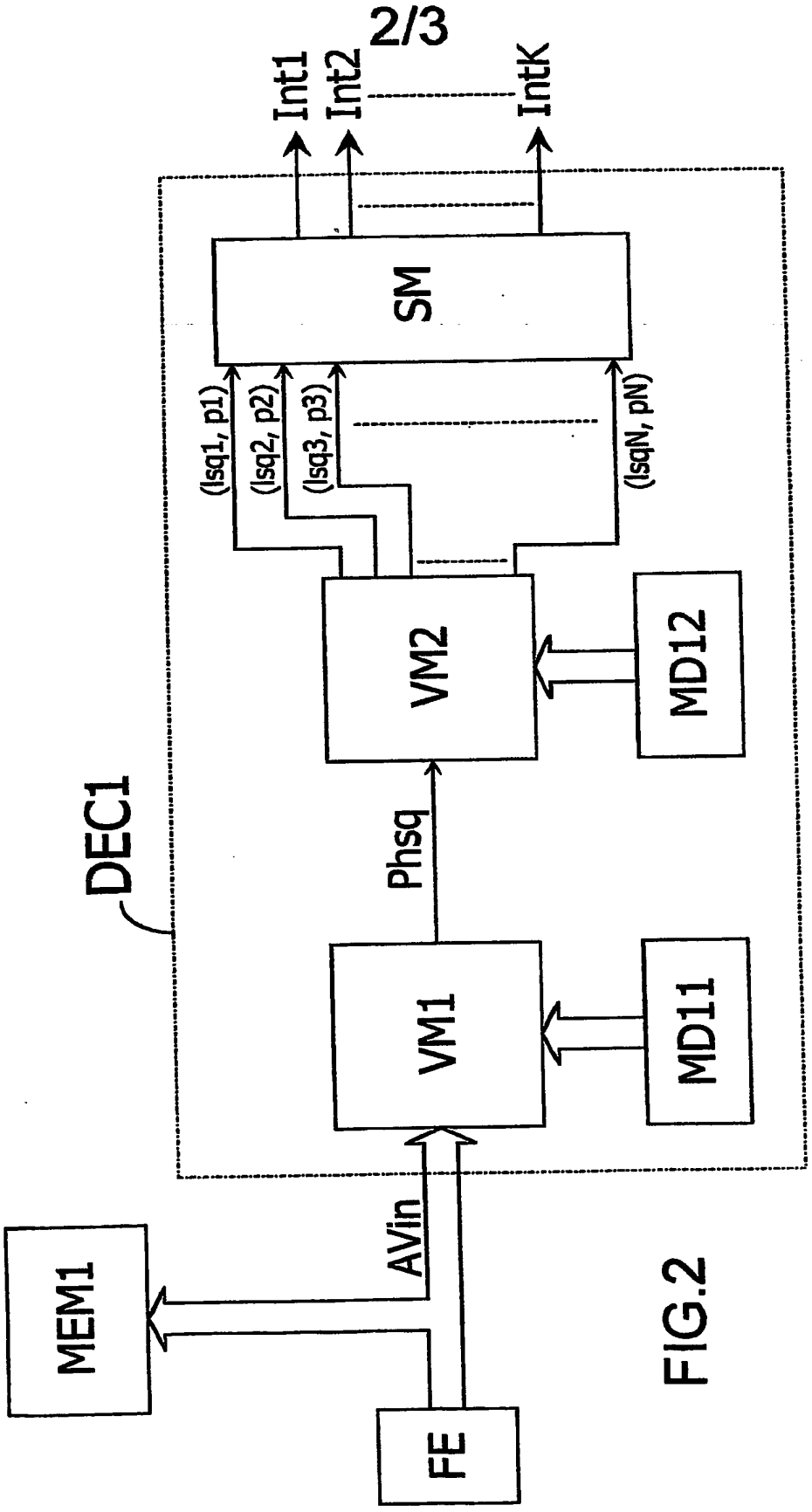

Int1
Int2
IntK
SM
(lsq1, p1)
(lsq2, p2)
(lsq3, p3)
(lsqN, pN)
VM2
MD12
DEC1
Phsq
VM1
MD11
AVin
MEM1
FE

FIG.2

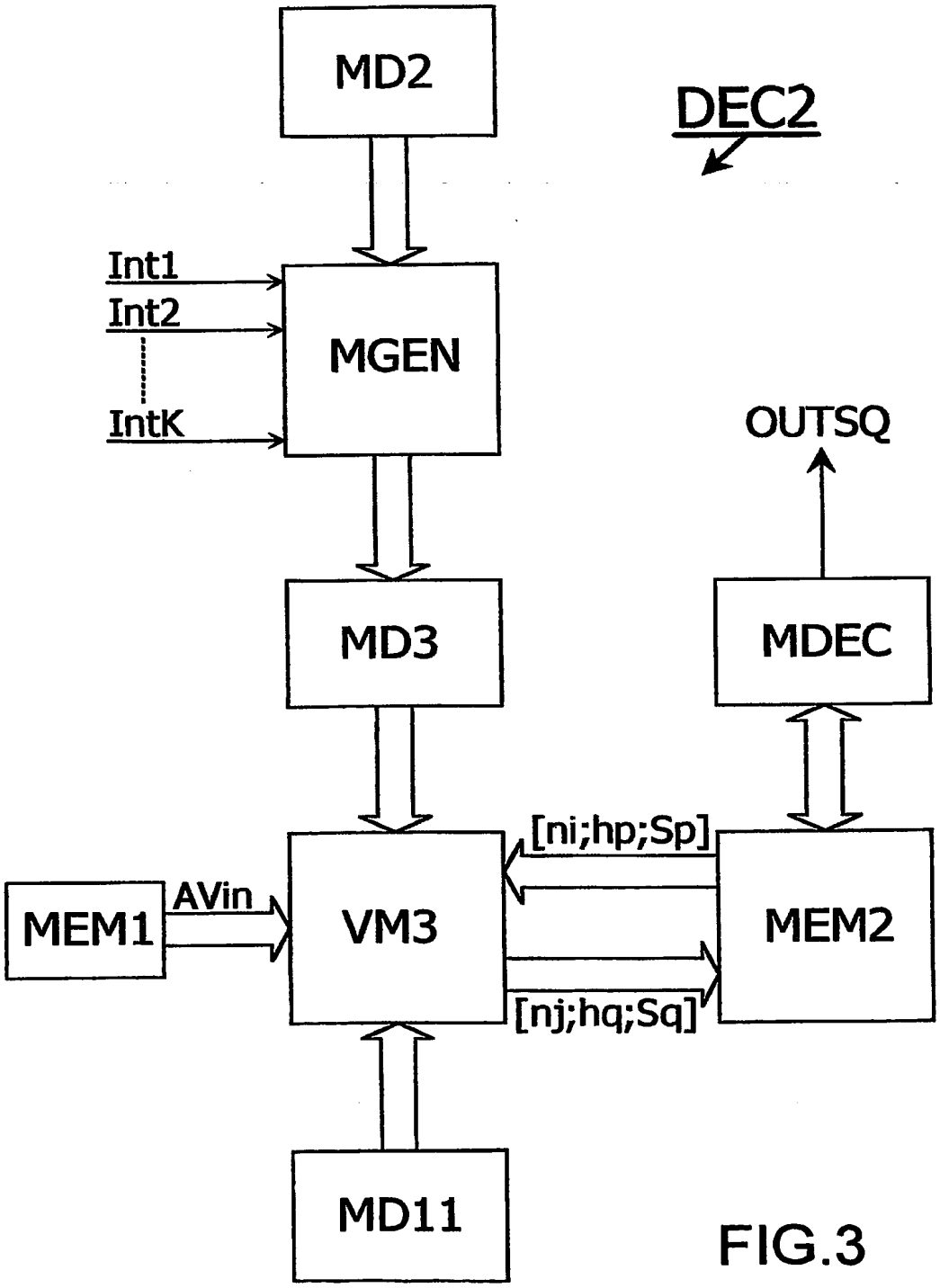

MD2
DEC2
Int1
Int2
IntK
MGEN
OUTSQ
MD3
MDEC
[ni;hp;Sp]
MEM1
AVin
VM3
MEM2
[nj;hq;Sq]
MD11

FIG.3

INTERNATIONAL SEARCH REPORT

International Application No
PCT/FR 03/00653

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G10L15/14 G10L15/18

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 715 298 A (IBM) 5 June 1996 (1996-06-05) abstract; figures 6,7 page 2, line 25-39	1-6

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 July 2003	21/07/2003

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Quélavoine, R

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/FR 03/00653

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0715298 A	05-06-1996	US 5729656 A	17-03-1998
		DE 69518723 D1	12-10-2000
		DE 69518723 T2	23-05-2001
		EP 0715298 A1	05-06-1996

RAPPORT DE RECHERCHE INTERNATIONALE

Demande Internationale No
PCT/FR 03/00653

A. CLASSEMENT DE L'OBJET DE LA DEMANDE
CIB 7 G10L15/14 G10L15/18

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)
CIB 7 G10L

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si réalisable, termes de recherche utilisés)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie °	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	EP 0 715 298 A (IBM) 5 juin 1996 (1996-06-05) abrégé; figures 6,7 page 2, ligne 25-39	1-6

☐ Voir la suite du cadre C pour la fin de la liste des documents

☒ Les documents de familles de brevets sont indiqués en annexe

° Catégories spéciales de documents cités:

"A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent

"E" document antérieur, mais publié à la date de dépôt international ou après cette date

"L" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)

"O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens

"P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

"T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention

"X" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

"Y" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

"&" document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée	Date d'expédition du présent rapport de recherche internationale
10 juillet 2003	21/07/2003

Nom et adresse postale de l'administration chargée de la recherche internationale	Fonctionnaire autorisé
Office Européen des Brevets, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Quélavoine, R

RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Demande Internationale No
PCT/FR 03/00653

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
EP 0715298 A	05-06-1996	US 5729656 A	17-03-1998
		DE 69518723 D1	12-10-2000
		DE 69518723 T2	23-05-2001
		EP 0715298 A1	05-06-1996